(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/13138 A1

(51) International Patent Classification[7]: G06T 1/00

(21) International Application Number: PCT/IB00/01089

(22) International Filing Date: 3 August 2000 (03.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **DIGITAL COPYRIGHT TECHNOLOGIES AG** [CH/CH]; Stauffacherstrasse 149, CH-8004 Zürich (CH).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **VOLOSHYNOVSKIY, Svyatoslav** [UA/CH]; Boulevard du Pont D'arve 6, CH-1205 Geneva (CH). **DEGUILLAUME, Frederic** [FR/CH]; 12, promenade Jean Treina, CH-1203 Geneva (CH). **HERRIGEL, Alexander** [DE/CH]; Bergstrasse 62, CH-8702 Meilen (CH). **PUN, Thierry** [FR/CH]; 60, chemin de la Gradelle, CH-1224 Chêne-Bougeries (CH).

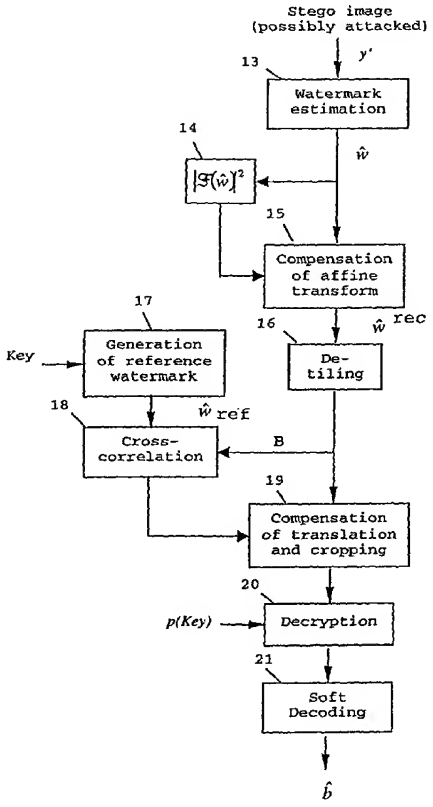
(74) Agent: **E. BLUM & CO.**; Vorderberg 11, CH-8044 Zürich (CH).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD FOR ADAPTIVE DIGITAL WATERMARKING ROBUST AGAINST GEOMETRIC TRANSFORMS



(57) Abstract: A method for digital content adaptive watermarking robust against general affine transforms, cropping and compression is disclosed. The method is based on a wavelet domain additive watermarking with a multiresolution perceptual mask (11) determined by the stochastic noise visibility function NVF of the cover image x. It is shown how to encode messages b and how to design the periodic watermark w in order to recover, based on the watermark Fourier magnitude spectrum F(w), general affine transform and compression attacks. Furthermore, it is demonstrated that the method is flexible and compatible with any message encoding technique and in particular with turbo codes, BJCR-, log-MAP and max-log-MAP decoders and with low-density parity check codes.

WO 02/13138 A1

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Method for Adaptive Digital Watermarking Robust Against Geometric Transforms

TECHNICAL FIELD

The invention refers to the field of digital watermarking and in particular to generating and extracting digital watermarks for images or video sequences.

BACKGROUND ART

Two major conflicting constraints on digital image watermarks are invisibility, i.e. avoiding perceptible artifacts in the watermarked or stego image, and robustness, i.e. resistance against various intentional or unintentional attacks such as affine geometric transforms (rotation, scaling, aspect ratio changes, shear), translation, cropping, image compression etc.

In earlier solutions the information to be embedded was encoded using e.g. M-ary modulation (M. Kutter, "Performance Improvement of Spread-Spectrum based Image Watermarking Schemes through M-ary Modulation", Lecture Notes in Computer Science: Third International Workshop on Information Hiding, Springer, Vol. 1768, 237-252) or algebraic error correction codes (ECC) (J. R. Hernandez, F. Perez-Gonzalez, J. M. Rodrigez and G. Nieto, "The impact of channel coding on the performance of spatial watermarking for copyright protection", Proc. ICASSP'98, 2973-2976, May 1998). M-ary encoding suffers from a high complexity of the watermark demodulator, whereas error correction codes are less effective. On the other hand, turbo codes and BCJR, log-MAP or max-log-MAP decoders (C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes", IEEE Trans. Comm., 1261-1271, October 1996) or low-density parity

check codes (R. Gallager, "Low-density parity-check codes", IRE Transactions on Information Theory, January 1962) have not been applied to digital watermarking.

The perceptual mask has to determine the optimal level of allowable distortions for the watermark embedding. An overview of empirical masking methods based on the deterministic models of the human visual system (HVS) is given by S. Voloshynovskiy, A. Herrigel, N. Baumgärtner and T. Pun, "A Stochastic Approach to Content Adaptive Digital Image Watermarking", Lecture Notes in Computer Science: Third International Workshop on Information Hiding, Springer, Vol. 1768, 211-236. The main problem consists in the content-adaptive watermarking, since in the most cases the HVS mask is given in the coordinate domain and watermark embedding is performed in some transform domain (block-wise and full-frame discrete Fourier (DFT) or discrete cosine (DCT) transforms, wavelet or Radon transforms). The embedded watermark is then transformed to the coordinate domain and mapped by the mask. More recent methods try to utilize either transform domain masking based on a just noticeable difference that originates from the image compression applications (I. Podilchuk and W. Zeng, "Image-Adaptive Watermarking Using Visual Models", IEEE Journal on Selected Areas in Communications, **16**(4), 525-539), or combined masking in frequency and coordinate domains (U. S. Pat. No. 6,031,914). In the latter, a major drawback is that both a frequency-domain and a spatial-domain perceptual mask must be applied consecutively in order to achieve invisibility. Furthermore, the watermark can only be extracted when the unmarked image is accessible.

In the above-mentioned publication by S. Voloshynovskiy et al. a stochastic perceptual mask based on a noise visibility function NVF is proposed. However, since the NVF and the perceptual mask are developed only in the spatial coordinate domain, they are not well adapted for calcula-

tions in a frequency domain and are not easily modifiable by restrictions stemming from the frequency domain.

Robustness against geometrical distortions has so far been relied on using a transform invariant domain (J. Oruanaidh and T. Pun, "Rotation, Scale and Translation Invariant Spread Spectrum Digital Image Watermarking", Signal Processing **66**(3), 303–317, 1998), or an additional template (WO 96/36163), or an Autocorrelation Function (ACF) of the watermark itself (M. Kutter, "Watermarking resistent to translation, rotation and scaling", Proc. SPIE Int. Symp. on Voice, Video, and Data Communication, 1998). The transform invariant domain approach suffers from interpolation and accuracy problems, therefore requires comparatively large images of size 512x512, and cannot recover rotational and aspect ratio changes simultaneously. The template approach needs a computationally expensive exhaustive search for recovering these transforms together, and it is susceptible to unauthorized removal of template peaks. In the ACF approach the watermark is replicated in the image in order to create 4 repetitions of the same watermark. The corresponding 9 peaks in the ACF are used to recover undergone geometrical transformations. However, the descending heights of the ACF peaks shaped by the triangular envelope function reduce the robustness of this approach against geometrical attacks accompanied by a lossy compression. The need for computing two discrete Fourier transforms (DFT) of double image size to estimate the ACF poses problems in real time applications with large images.

A further requirement for digital watermarking is a sufficient information capacity of the watermark. In order to attach a unique identifier to each buyer of an image, a typical watermark should be able to carry at least 60-100 bits of information. However few publications deal with 60 or more bits.

From the above review it is concluded that the existing technologies exhibit at least one of the following problems:

1. Constrained spatial domain modulation for content-adaptive watermarking.

2. Inability to resist against geometrical transforms accompanied by the lossy JPEG compression.

3. Low simultaneous robustness against lossy JPEG (DCT-based) and wavelet compression.

4. Low robustness against printing/rescanning for high quality commercial magazine printing.

5. No protection against intentional template removal.

6. Less than 60 bits encoding for limiting the complexity of the watermark demodulator or decoder.

DISCLOSURE OF THE INVENTION

It is the object of the present invention to provide an improved method of the type mentioned above that is in particular capable of dealing with at least some, preferably all of these problems. This object is achieved by the subject-matter as set forth in the independent claims. Preferred embodiments are described in the dependent claims. The present invention is well suited for watermarking still images and video data.

The invention resides in a method for embedding a digital watermark w in an image x, comprising the steps of encoding a digital message b in a codeword c, mapping the codeword c and allocating the mapped codeword c into a block B, producing a symmetric block B' of fourfold size by flipping and copying the block B once in every block direction, tiling the symmetric block B' in order to generate a symmetric digital watermark w with a period B'

and embedding the watermark w in the image x in order to obtain a stego image y. By tiling or repeating the basic block B' a plurality of times, periodic features are introduced into the final watermark w both in the coordinate and frequency domain, that can be used for retrieving affine transform attacks undergone by the stego image. The block flipping makes the watermark w robust against stego image flipping attacks, i.e. rotations by 90°, 180° or 270°, and reduces the number of ambiguities during estimation of the undergone geometrical attacks. Furthermore, the block flipping increases the invisibility of the watermark w by visually decorrelating its repetitive structure in the coordinate domain.

Preferred embodiments are: adding a secret-key-dependent reference watermark w_{ref} in remaining orthogonal spatial locations of the block B to render the resulting watermark w robust against translation or cropping attacks undergone by the stego image y; up-sampling pixels of the block B or equivalently B' at least twofold in each block dimension for creating robustness against the finite resolution of image input or output media, such as printers and scanners; using a turbo code or a low-density parity check code for encoding the digital message b thereby keeping the block size small; using a secret encryption key for encrypting the codeword c and/or a secret block allocation key for block allocation to improve the safety of message hiding and decoding; embedding the watermark w in the image x in wavelet sub-bands k,l, wherein k is a resolution index and l a direction index thereby providing full compatibility of the embedding procedure with the recently developed wavelet-based compression standard JPEG2000.

The invention further resides in a method for embedding a watermark w in an image x, comprising the steps of: calculating image wavelet components $\tilde{x}_{k,l}(i,j)$ and watermark wavelet components $\tilde{w}_{k,l}(i,j)$ for pixel locations i,j, based

on the $\tilde{x}_{k,l}(i,j)$ calculating in the wavelet sub-bands k,l a noise visibility function $NVF_{k,l}(i,j)$ and therefrom a perceptual mask $PM_{k,l}(i,j)$ for masking the $\tilde{w}_{k,l}(i,j)$ and embedding the masked watermark wavelet components into the $\tilde{x}_{k,l}(i,j)$ to produce stego image wavelet components $\tilde{y}_{k,l}(i,j)$ and calculating by an inverse discrete wavelet transformation (IDWT) the stego image y. By using a stochastic approach to image analysis based on the NVF and by defining in the wavelet domain the NVF and a NVF-based perceptual mask PM, invisibility constraints, frequency-domain constraints and geometric robustness constraints can be incorporated into a single perceptual mask PM.

Preferred embodiments refer to: calculating the noise visibility function $NVF_{k,l}(i,j)$ from a stationary generalized Gaussian model or a non-stationary Gaussian model of the image x; incorporating in the perceptual mask $PM_{k,l}(i,j)$ watermark strengths $S^e_{k,l}$ for edges and textures of the image x with a weighting factor 1-NVF and watermark strengths $S^f_{k,l}$ for flat regions of the image x with a weighting factor NVF; wavelet-domain embedding by multiplying $PM_{k,l}(i,j)$ with $\tilde{w}_{k,l}(i,j)$ and adding $\tilde{x}_{k,l}(i,j)$; adapting the watermark strengths $S^e_{k,l}$ and/or $S^f_{k,l}$ in order to take advantage of a frequency-dependent modulation transfer function (MTF) and/or a spatial orientational dependence of the human visual system (HVS); in particular choosing $S^e_{k,l} \geq S^f_{k,l}$ for a majority of or all wavelet sub-band indices k, l and/or choosing $S^e_{1,1}>S^e_{2,1}>S^e_{3,1}>S^e_{4,1}<S^e_{5,1}$ and $S^f_{1,1}>S^f_{2,1}>S^f_{3,1}>S^f_{4,1}<S^f_{5,1}$ for k=1...5 and/or choosing $S^e_{k,1}\leq S^e_{k,3}$, $S^e_{k,2}\leq S^e_{k,3}$ and $S^f_{k,1}\leq S^f_{k,3}$, $S^f_{k,2}\leq S^f_{k,3}$, wherein the indices l=1 and l=2 denote a horizontal and vertical orientation and l=3 a diagonal orientation in the image x; and/or compressing the image x in the wavelet sub-bands k, l before the watermark embedding in order to realize "compressed domain watermarking".

The invention further resides in a method for extracting a watermark w, that was previously embedded according to

invention, from a possibly attacked stego image y', comprising the steps of: calculating an estimated watermark $\hat{w}$ from the stego image y', detiling the estimated watermark $\hat{w}$ into the symmetric block B' by summing corresponding portions of a plurality of periods of the estimated watermark $\hat{w}$ and converting the symmetric block B' into the block B and extracting the digital message b from the block B. This extraction method assures that full advantage is taken of the tiling and flipping operations performed during watermark embedding.

Preferred embodiments refer to: using a maximum a posteriori probability (MAP estimation) for calculating the estimated watermark $\hat{w}$; estimating a watermark-covariance matrix R_w globally; estimating an image-covariance matrix $\hat{R}_x$ locally; estimating and correcting a geometric affine transform from peaks in the spectral power density $|F(\hat{w})|^2$ and/or the autocorrelation function (ACF) $\hat{w}*\hat{w}$ of the estimated watermark $\hat{w}$; cross-correlating the block B with a reference watermark w_{ref} to compensate translations and/or cropping; down-sampling a previously up-sampled block B by averaging identical neighbouring pixels; using secret key for block deallocation and/or message decryption; and/or using a BJCR, a log-MAP or a max-log-MAP decoder for soft decoding previously turbo-coded digital messages b.

Other objects, features and advantages of the present invention will become apparent from the description in connection with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The drawings show in

Fig. 1 an embodiment for generating a digital watermark w robust against geometrical transforms;

Fig. 2a exemplary wavelet pyramids of a cover image x ("Lena"), in Fig. 2b of the digital watermark w, and in Fig. 2c of the noise visibility function NFV of the cover image x;

Fig. 3a the modulation transfer function (MTF) of the human visual system (HVS) and a state-of-the-art non-adaptive embedding;

Fig. 3b a 1-dimensional wavelet decomposition and Fig. 3c an adaptive embedding according to a preferred embodiment;

Fig. 4 a 2-dimensional wavelet decomposition related to the MTF;

Fig. 5 an embodiment for embedding the digital watermark w robustly in the wavelet domain;

Fig. 6 an embodiment for extracting and decoding the digital watermark w from an attacked stego image y'; and

Fig. 7a-7d watermark extraction using spectral power density peaks: watermark w=cover image x-stego image (Fig. 7a), an estimated

Fig. 7a-7d digital watermarks w, $\hat{w}$ extracted by using spectral power density peaks: cover image x-stego image y (Fig. 7a); watermark estimated by denoising a stego image y (Fig. 7b), a compressed stego image y' (Fig. 7c) and a rotated and compressed stego image y' (Fig. 7d).

In the drawings identical parts are designated by identical reference numerals.

MODES FOR CARRYING OUT THE INVENTION

Formulation of a preferred embodiment:

We formulate the embedding process as an additive content-adaptive watermarking in the wavelet domain with the watermark possessing special spatial structure that enables to recover general affine transforms. We assume that we are given a cover image to be watermarked denoted x. If it is an RGB image we work with the luminance component, though the same methodology can be applied to other color spaces. The given message (the copyright information or URL address) in binary form $b=(b_1,...,b_L)^T$ is to be embedded in the cover image $x=(x_1,...,x_N)^T$ of size $M_1 \times M_2$, where $N=M_1 \cdot M_2$.

Message encoding and spatial allocation:

Fig. 1 shows an example of watermark creation. The message b is first encoded 1 in a codeword c using preferably either low-density parity check codes (R. Gallager) or turbo codes (C. Berrou and A. Glavieux), the publications of which are herewith incorporated in this application in their entirety by reference. The maximum rate at which these codes can be used is known to be bounded below channel capacity. However, the existence of simple iterative decoding schemes and their outstanding error performance more than compensate this weakness.

The codeword c is then mapped 2 from {0,1} to {-1,1} and encrypted 3 by multiplying on a key-dependent sequence p with following spreading 4 over a square block B of size $N_1 \times N_1$ with some density D using a secret key. In the general case, it could also be a rectangular block B or a block B of any shape.

The key-dependent reference watermark w_{ref} is added 5 to the above block B in some or all remaining orthogonal spatial locations. The reference watermark w_{ref} is used to recover cropping and translation based on the cross-

correlation with the attacked stego image y'. The reference watermark w_{ref} consists of a binary key-dependent sequence {-1,1} and its length is determined by the embedding density (1-*D*) as is described above.

The resulting block B is up-sampled 6 by the factor 2 to receive a low-pass watermark and then flipped and copied 7 once in each direction, producing a symmetric block B' of size $4N_1 \times 4N_1$. The flipping 7 is performed to visually decorrelate the structure of the repeated watermark w and to reduce the number of ambiguities during estimation of the undergone geometrical attacks. Finally, the $4N_1 \times 4N_1$ block B' is repeated preferably over the whole image size, resulting in a symmetrical and periodical watermark w with periods $T_1 = T_2 = 4N_1$. In our applications we use *L*=64 bit messages that are encoded using the turbo code (*K*=132). The block size is chosen to be N_1=19 resulting in a density *D*=0.74 in order to have exactly 2 times repetition of the codeword c in every block B. The final block B' after up-sampling 6 and flipping 7 has the size 76×76. The scheme is very flexible in respect to the encoding 1 and can use any known modulation technique or even more advanced error correction codes (ECC).

Stochastic multi-resolution image modeling and watermark embedding:

The principle of watermark embedding is shown in Fig. 5. To embed the above obtained watermark w in a cover image x a linear additive scheme is used in the wavelet domain. Both the cover image x and the watermark w are first decomposed into multi-resolution sub-band pyramids using the (discrete) Forward Wavelet Transform (FWT or DWT). First, the cover image x is padded to a square size of the nearest larger number to the original cover image size of power of 2 in order to apply a standard wavelet transform DWT, 9. In the numeric example below, $N_w = 5$ levels are used for the DWT based on the Daubechies 8-tap filter (M. Vetterli and J. Kovacevic, "Wavelets and Sub-

band Coding", Prentice Hall, 1995). This results in 6 resolution sub-bands k or scales. Scales from 1 to $N_w = 5$ are also divided into 3 components corresponding to distinct orientations l, for horizontal (H), vertical (V) and diagonal (D) directions. The lowest scale $k = N_w + 1$ however consists of only a low-pass component. Fig. 2a shows the pyramids of the cover image x and Fig. 2b of the watermark w.

The watermarking process is applied and adapted to each (k,l) wavelet sub-band component separately as described below. Finally, the stego image y is reconstructed by computing the Inverse Wavelet Transform (IWT, 12) of the digitally watermarked image pyramid.

An important issue is the adaptation of the watermark w to the properties of the HVS, i.e. content-adaptive watermarking. Assuming we are given a masking function of the HVS, we wish to embed the above described watermark into the cover image x keeping it under the threshold of visual imperceptibility. We propose to use a stochastic perceptual mask $PM_{k,l}(i,j)$, 11 based on a noise visibility function (NVF) proposed by Voloshynovskiy et al and earlier developed only for the coordinate domain. Here the NVF is for the first modified in order to include the multi-resolution paradigm in the stochastic framework to take into account a modulation transfer function (MTF) of the HVS and to match the proposed watermarking algorithm with the recently developed image compression standard JPEG2000 for future integration. This practically means that different watermark strengths S or S^e, S^f are assigned to different image sub-bands k, l. Such a modification leads to a non-white spectrum of watermarks w being matched with the MTF. Previously this could not be achieved with the coordinate-domain based version of the NVF. The second reason to use wavelet domain embedding is motivated by the desire to incorporate the anisotropy of the HVS to different spatial directions in the perceptual

mask $PM_{k,l}(i,j)$, 11. The coordinate domain version of the NVF used only an isotropic image decomposition based on the extraction of a local mean from the original image or its high-pass filtered counterpart. In the wavelet domain k, l the image coefficients in 3 basic spatial directions, i.e. horizontal (l=1), vertical (l=2) and diagonal (l=3), are received as a result of the decomposition, which therefore allows to exploit the anisotropic sensitivity of the HVS. As a result, the watermark strengths S can be varied for different orientations l in the proposed mask $PM_{k,l}(i,j)$, 11.

The NVF is based on a stationary Generalized Gaussian (sGG) model or on a non-stationary Gaussian model of the cover image x or the cover image wavelet coefficients $\tilde{x}_{k,l}(i,j)$ for every sub-band k, l. Accordingly the perceptual edge and texture masking in the wavelet domain is modeled based on the NVF, of pixel (i,j), for each sub-band component (k,l):

$$NVF_{k,l}(i,j)=\frac{\tilde{\omega}_{k,l}(i,j)}{\tilde{\omega}_{k,l}(i,j)+\sigma^2_{\tilde{x}_{k,l}}}, \qquad \text{(G1)}$$

$\sigma^2_{\tilde{x}_{k,l}}$ is the global variance of the wavelet image coefficients from sub-band (k,l), and the watermark wavelet components $\tilde{w}_{k,l}(i,j)$ can be written as

$$\tilde{\omega}(i,j)=\gamma_{k,l}\left[\eta(\gamma_{k,l})\right]^{\gamma_{k,l}}\frac{1}{\left|r_{k,l}(i,j)\right|^{2-\gamma_{k,l}}} \quad \text{with} \qquad \text{(G2)}$$

$$\eta(\gamma)=\sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}}, \quad \Gamma(t)=\int_0^\infty e^{-u}u^{t-1}du \quad \text{and} \quad r_{k,l}(i,j)=\frac{\tilde{x}_{k,l}(i,j)}{\sigma_{\tilde{x}_{k,l}}} \qquad \text{(G3)}$$

where $\Gamma(t)$ is the gamma function. The NVF's features for a given sub-band k, l are determined by the global sub-band variance $\sigma^2_{\tilde{x}_{k,l}}$ and by the shape parameter $\gamma_{k,l}(i,j)$ which is estimated based on the moment matching method (A. Jain, "Fundamentals of digital image processing",

Prentice-Hall, 1989). An example of the NVF pyramid for image "Lena" is shown in Fig. 2c.

Finally the weighted watermark is added to the cover image x using the following embedding rule:

$$\tilde{y}_{k,l}(i,j) = \tilde{x}_{k,l}(i,j) + \left(\left(1 - NVF_{k,l}(i,j)\right)\cdot S^{e}_{k,l} + NVF_{k,l}(i,j)\cdot S^{f}_{k,l}\right)\cdot \tilde{w}_{k,l}(i,j) \qquad \text{(G4)}$$

wherein the factor in front of the $\tilde{w}_{k,l}(i,j)$ defines the perceptual mask $PM_{k,l}(i,j)$. The $\tilde{y}_{k,l}(i,j)$ are the obtained stego wavelet components and $PM_{k,l}(i,j) \cdot \tilde{w}_{k,l}(i,j)$ are the perceptually masked watermark wavelet components. $S^{e}_{k,l}$ is an embedding strength for the edges and textures, and $S^{f}_{k,l}$ is a strength for the flat regions of the cover image x. Visual masking is ensured first by choosing $S^{e}_{k,l}$ greater than $S^{f}_{k,l}$ for edges and textures hiding, and second by using adapted strengths for each resolution, and even for each orientation based on the properties of the MTF. An example of practically used embedding parameters according to the MTF properties, considering cover image pixel values in the range [0,255], are:

$$S^{e}_{k,l} = \begin{vmatrix} 18 & 18 & 20 & 0 \\ 11 & 11 & 15 & 0 \\ 5 & 5 & 7 & 0 \\ 2 & 2 & 4 & 0 \\ 5 & 5 & 7 & 1 \end{vmatrix} \qquad S^{f}_{k,l} = \begin{vmatrix} 0.1 & 0.1 & 0.2 & 0 \\ 0.2 & 0.2 & 0.5 & 0 \\ 0.5 & 0.5 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 2 & 2 & 3 & 1 \end{vmatrix}$$

where rows k denote decreasing resolutions, and columns l each orientation. The watermark strengths or embedding parameters $S^{e}_{k,l}$, $S^{f}_{k,l}$ reflect very important particularities of the HVS. First, the strengths of watermark for the diagonal directions are chosen to be higher than for the vertical and horizontal ones. This is motivated by the fact that the anisotropy sensitivity of the HVS to the diagonally oriented patterns is lower than for the vertical and horizontal directions. Therefore, it makes possible to embed stronger watermark signals there. Moreover, it allows to obtain, as a result, better robustness against lossy compression (both JPEG-DCT and wavelet JPEG2000). The lossy compression is exploiting the same

property of the HVS to allocate smaller amounts of bits in the diagonal directions for the image coding. Therefore, the proposed embedding technique utilizes both information about the HVS and the quantization of lossy image coding to increase the robustness of the watermark w.

Second, the MTF of the HVS has a typical frequency dependence, as is shown in Fig. 3a (A. Jain, p. 55), with a maximum in a low frequency range and decreasing side lobes at very low and middle to high frequencies. In the case of non-adaptive watermark embedding (Fig. 3a), the typical additive white Gaussian watermark has a uniform spectrum. A uniform increase of the watermark strength or equivalently watermark power density would violate the invisibility constraint at low frequencies. However, there still remains a lot of space for watermark embedding at the very low, middle and high frequencies below the threshold of imperceptibility. To exploit this opportunity we use the wavelet sub-band decomposition (Fig. 3b: wavelet subbands V1…V5 for a 1-dimensional example), wherein the watermark strength could be adopted according to the local properties of the MTF (Fig. 3c). This adaptation to the MTF is reflected in the proper choice of the embedding parameters S^e, S^f that have maxima in the corresponding frequency sub-bands k along each spatial direction l (Fig. 4).

Third, the particular properties of the given image x within each sub-band k, l are taken into account using local watermark strength control based on the NVF, as discussed earlier. This feature has image dependent character oppositely to the previous two properties that characterize the HVS in general. Therefore, the proposed watermark embedding technique utilizes both general features of the HVS as well as local statistics of cover images x.

Watermark extraction and message decoding:

A generalized block-diagram of watermark extraction is shown in Fig. 6. The embedded watermark w is first estimated 13, $\hat{w}$ from the possibly attacked stego image y'. Secondly, geometric distortions which may have occurred are retrieved and compensated 15 to obtain a rectified watermark $\hat{w}^{rec}$, by analyzing 14 the Fourier transform $F(\hat{w})$ or the spectral power density magnitude $|F(\hat{w})|^2$ and/or an autocorrelation function (ACF) $\hat{w}*\hat{w}$ of the estimated watermark $\hat{w}$. The ACF is preferably obtained by $\hat{w}*\hat{w}= F^{-1}(|F(\hat{w})|^2)$ with $F^{-1}()$ being the inverse DFT. The tiled blocks are then detiled or averaged 16 in order to get an estimate of the embedded redundant sequence according to the maximum likelihood (ML) estimate for a Gaussian channel. The cropping and translation are compensated 19 using cross-correlation 18 with the reference key-dependent watermark w_{ref}, 17. Finally, the message is decrypted 20 and decoded 21.

Watermark estimation:

To estimate the watermark w a maximum a posteriori probability (MAP) estimate is used:

$$\hat{w} = \arg\max_{\tilde{w}\in\Re^N}\{p_X(y'|\tilde{w})\cdot p_W(\tilde{w})\}, \qquad \text{(G5)}$$

wherein $p_W(.)$ is the probability density function of the watermark w. Assuming that the image y' and watermark w are conditionally independent identically distributed locally Gaussian, i.e. $x \sim N(\bar{x}, R_x)$ and $w \sim N(0, R_w)$ with the covariance matrices R_x of the image x and R_w of the watermark w, where R_w also includes the effect of perceptual watermark modulation, one can determine:

$$\hat{w} = \frac{R_w}{R_w + \hat{R}_x}(y' - \bar{y}') \qquad \text{(G6)}$$

where the mean values $\bar{y}' \approx \bar{x}$ are assumed to be equal and where $\hat{R}_x = \max(0, \hat{R}_y - R_w)$ is the ML-estimate of the local

variance ($\hat{R}_x = \sigma_x^2 I$ with I=identity matrix) and $\hat{R}_y$ is an estimated covariance matrix of the original stego image y.

An important issue is the estimation of the watermark covariance matrix R_w in the above estimate. This can be done based on the available copy of the stego image y'. However, the severe distortions due to lossy JPEG compression could destroy the information about the texture masking that was used for the watermark embedding, and a histogram modification attack could damage the relevant information about contrast sensitivity masking. Since no reliable information about the perceptual mask PM is available after these attacks, we propose to use a global estimate of the watermark strength based on the available copy of the attacked image y'. This practically means that we assume spatial stationarity of the watermark $\hat{R}_w = \hat{\sigma}_w^2 I$. To estimate a global watermark variance we use the following formula:

$$\hat{R}_w = \frac{1}{NM} \sum_{m=1}^{N} \sum_{n=1}^{M} \hat{\sigma}_y^2(m,n) \qquad \text{(G7)}$$

where $\hat{\sigma}_y^2(m,n)$ is a local variance of the stego image y in the coordinates (m,n), for an image of size $N \times M$. The estimate (G7) is a global mean value of the watermark variance. Obviously, other robust versions of (G7) such as a robust median estimate of the variance could be applied, as well.

Determining affine geometrical distortions:

To determine the affine transformation applied to the image we compute $|F(\hat{w})|^2$ from the estimated watermark $\hat{w}$, where $F(\hat{w})$ is the discrete FT. Due to the periodicity of the embedded information, the estimated watermark spectrum possesses a discrete structure. Assuming that the watermark w is white noise within the block B, the spectrum of the watermark w will additionally be uniform. Therefore, $|F(\hat{w})|^2$ shows aligned and regularly spaced peaks. For a $T_1 \times T_2$-periodical watermark $\hat{w}$, peaks will

have periods M_1/T_1 and M_2/T_2 for a 2-D FT domain of size $M_1 \times M_2$. If an affine distortion was applied to the stego image y, the peaks layout will be re-scaled, rotated and/or sheared, but alignments will be preserved. Therefore, it any affine geometrical distortion can be estimated from these peaks by fitting alignments and estimating periods.

Finding the matched points between the extracted positions of local peaks in the magnitude spectrum of the estimated watermark (z_1, z_2) and the reference grid (f_1, f_2) computed based on the knowledge of the embedded watermark period, one can estimate the linear affine transform A using all matched points such that the next criterion is minimized:

$$\Phi = \min_A \rho\left\{ A \begin{bmatrix} f_1 f_2 \\ \vdots \\ f_k f_k \end{bmatrix}^T - \begin{bmatrix} z_1 z_2 \\ \vdots \\ z_k z_k \end{bmatrix}^T \right\} \qquad \text{(G8)}$$

where $\rho\{\}$ is a negative log-likelihood function associated with the distribution of the misaliments and k is a number of matched points. In the most common case, it is assumed that the misalignment distribution is Gaussian, and one receives a quadratic log-likelihood function $\rho\{\}=\|\cdot\|^2$ and the corresponding mean square error minimization criterion. In the more general case, the above problem could be solved based on the theory of robust M-estimators, if some ambiguity about misalignment distribution exists.

Fig. 7a-7d show peaks extracted from the magnitude spectrum of the watermark $|F(\hat{w})|^2$. In Fig. 7a, the real embedded watermark $\hat{w}$ is shown that was calculated by subtracting y-x using the knowledge of the cover image x in a non-oblivious approach, whereas in Fig. 7b the Wiener predicted watermark $\hat{w}$ is taken. Therefore, these peaks can be extracted from the stego data with high fidelity

based on the estimated watermark $\hat{w}$ without knowledge of the cover image x. This important conclusion is also connected with the observation that the embedded watermark w is mostly allocated in the middle frequency band. This has double importance. First, low frequencies of the stego image y or y' are not altered considerably in order not to produce visible distortions. Second, the watermark w will resist against lossy compression that removes mostly high frequency components from the image y or y'.

Fig. 7c-7d show peaks extracted after lossy compression, without (Fig. 7c) and with (Fig. 7d) geometric distortions, here a 37° rotation of the stego image y' followed by a JPEG compression with a quality factor QF=50%. In experiments peaks could be properly extracted from JPEG compressed images with a quality factor QF up to 50. At the time of patent submission, no known watermarking method is able to resist to affine transforms combined with such a compression.

Recovering translation and cropping is based on the reference key-dependent watermark w_{ref}, 17 (Fig. 6). To reduce computational complexity and using the information about the periodicity of the watermark $\hat{w}$ we first perform watermark detiling 16, i.e. coherent summation of the estimated watermark $\hat{w}$ from different periods. This results in the symmetric block B' that is converted to the final block B of size $N_1 \times N_1$. The block B is correlated 18 with the reference watermark w_{ref}. The maximum of cross-correlation 18 makes possible to establish the undergone translation or cropping that is easily compensated 19.

Message decoding:

Assuming that attack, prediction and extraction errors could be modeled as additive Gaussian, the detector is designed using the ML formulation for the detection of a known signal (projection sets p are known due to the key) in Gaussian noise, that results in a correlator detector

$$r = \langle \hat{w}, p \rangle \quad \text{(G9)}$$

In more general cases, the detector should be designed for stationary non-Gaussian noise or for the non-stationary Gaussian case. Finally, given an observation vector r, the decoder that minimizes the conditional probability of error, assuming that all codewords b are equi-probable, is given by the ML decoder:

$$\hat{b} = \arg\max_{\tilde{b}} p\left(r \mid \tilde{b}, x\right) \quad \text{(G10)}$$

Based on the central limit theorem (CLT) most researchers assume that the observed vector r can be accurately approximated as the output of an additive Gaussian channel noise for a large sample space.

We use symbol-by-symbol MAP decoder for the turbo code that is commonly known as a BCJR decoder, a log-MAP or a max-log-MAP decoder, i.e. soft decoding, that is known to be superior in comparison with the hard decoding for Gaussian channels.

While there are shown and described presently preferred embodiments of the invention, it is to be distinctly understood that the invention is not limited thereto but may be otherwise variously embodied and practiced within the scope of the following claims.

CLAIMS

1. A method for embedding a digital watermark w in an image x, comprising the steps of
   a) encoding (1) a digital message b in a codeword c,
   b) mapping (2) the codeword c and allocating (4) the mapped codeword c into a block B,
   c) producing (7) a symmetric block B' by flipping and copying the block B once in every block direction,
   d) tiling (8) the symmetric block B' in order to generate a periodic symmetric digital watermark w and
   e) embedding the watermark w in the image x in order to obtain a stego image y.

2. The method according to claim 1, comprising, between steps b) and c), the step or steps of
   a) adding (5) a secret-key-dependent reference watermark w_{ref} to the block B in remaining orthogonal spatial locations of the block B and/or
   b) up-sampling (6) pixels of the block B at least twofold in each block dimension.

3. The method according to one of the claims 1-2, comprising the steps of
   a) using a turbo code or a low-density parity check code for encoding (1) the digital message b and/or
   b) using a secret encryption key for encrypting (3) the codeword c and/or a secret block allocation key for block allocation (4) and/or
   f) embedding the watermark w in the image x in wavelet sub-bands (k,l).

4. A method for embedding a watermark w in an image x, in particular according to one of the previous claims, comprising the steps of
   a) calculating by a discrete wavelet transform (DWT, 9, 10) image wavelet components $\tilde{x}_{k,l}(i,j)$ of the image x and watermark wavelet components $\tilde{w}_{k,l}(i,j)$ of the

watermark w, for pixel locations (i,j) and wavelet sub-band indices k, 1,

b) based on the image wavelet components $\tilde{x}_{k,l}(i,j)$ calculating a noise visibility function $NVF_{k,1}(i,j)$ in the wavelet sub-bands (k,1) and therefrom a perceptual mask $PM_{k,1}(i,j)$ (11) for masking the watermark wavelet components $\tilde{w}_{k,l}(i,j)$ and

c) embedding the masked watermark wavelet components into the image wavelet components $\tilde{x}_{k,l}(i,j)$ to produce stego image wavelet components $\tilde{y}_{k,l}(i,j)$ and calculating by an inverse discrete wavelet transformation (IDWT, 12) the stego image y.

5. The method according to claim 4, comprising the steps of

a) calculating the noise visibility function $NVF_{k,1}(i,j)$ from a stationary generalized Gaussian model or a non-stationary Gaussian model of the image x and/or

b) based on the noise visibility function $NVF_{k,1}(i,j)$ calculating a perceptual mask (11)

$$PM_{k,1}(i,j) = (1-NVF_{k,1}(i,j)) \bullet S^{e}_{k,1} + NVF_{k,1}(i,j)) \bullet S^{f}_{k,1},$$

wherein $S^{e}_{k,1}$ are watermark strengths for edges and textures and $S^{f}_{k,1}$ are watermark strengths for flat regions of the image x, and/or

c) calculating the stego image wavelet components $\tilde{y}_{k,l}(i,j)$ according to an embedding rule

$$\tilde{y}_{k,l}(i,j) = \tilde{x}_{k,l}(i,j) + PM_{k,1}(i,j) \bullet \tilde{w}_{k,l}(i,j).$$

6. The method according to one of the claims 4-5, comprising the steps of watermark weigthing in the wavelet sub-bands (k,1) by watermark strengths $S^{e}_{k,1}$ for edges and textures and/or by watermark strengths $S^{f}_{k,1}$ for flat regions of the image x in order to exploit a frequency-dependent modulation transfer function (MTF) and/or a spatial orientational dependence of the human visual system (HVS).

7. The method according to claim 6, comprising the steps of

a) choosing the watermark strengths $S^e_{k,l}$ for edges and textures larger than the watermark strengths $S^f_{k,l}$ for flat regions for a majority of or all wavelet sub-band indices k, l and/or

b) adapting the watermark strengths $S^e_{k,l}$, $S^f_{k,l}$ as a function of the wavelet sub-band index k in an inverse relation to a modulation transfer function (MTF) of the human visual system (HVS) and/or

c) choosing the watermark strengths $S^e_{k,l}$, $S^f_{k,l}$ as a function of the wavelet sub-band index l larger for a diagonal orientation (l=3) than for a horizontal (l=1) or vertical (l=2) orientation.

8. The method according to one of the previous claims, comprising the step of subjecting the image x to a compression scheme in wavelet sub-bands k, l, in particular to JPEG2000 compression, before the embedding of the digital watermark w.

9. A method for extracting a watermark w from a possibly attacked stego image y', wherein an original stego image y was obtained by embedding the watermark w in an image x according to one of the claims 1-3 and in particular according to one of the claims 4-7, comprising the steps of

a) calculating (13) an estimated watermark $\hat{w}$ from the stego image y',

b) detiling (16) the estimated watermark $\hat{w}$ into the symmetric block B' by summing corresponding portions of a plurality of periods of the estimated watermark $\hat{w}$ and converting the symmetric block B' into the block B and

c) extracting (20, 21) the digital message b from the block B.

10. The method according to claim 9, comprising the steps of

a) using a maximum a posteriori probability (MAP estimation) for calculating the estimated watermark $\hat{w}$ and

b) in particular using an approximate equation

$$\hat{w} = R_w / (R_w + \hat{R}_x) \bullet (y' - \overline{y'}) ,$$

wherein R_w is a watermark-covariance matrix, $\hat{R}_x$ is an estimated image-covariance matrix, and $\overline{y'}$ are mean values of the stego image y'.

11. The method according to claim 10, comprising the steps of

a) estimating a watermark-covariance matrix R_w globally by averaging local variances $\sigma_{y'}^2(m,n)$ of the stego image y' over spatial coordinates (m,n) and/or

b) estimating an image-covariance matrix $\hat{R}_x$ locally from $\max(0, \hat{R}_y - R_w)$, wherein $\hat{R}_y$ is an estimated covariance matrix of the original stego image y based on a maximum likelihood estimate, R_w is a watermark-covariance matrix and max() defines a maximum of its arguments.

12. The method according to one of the claims 9-11, comprising the steps of

a) calculating (14) a spectral power density $|F(\hat{w})|^2$ of the estimated watermark $\hat{w}$, wherein $F(\hat{w})$ is a discrete Fourier transform (DFT), and/or calculating an autocorrelation function (ACF) $\hat{w} * \hat{w}$ of the estimated watermark $\hat{w}$,

b) extracting peaks from the spectral power density $|F(\hat{w})|^2$ and/or from the autocorrelation function (ACF) $\hat{w} * \hat{w}$,

c) based on the peaks estimating coefficients of a geometric affine transform matrix A and compensating (15) geometrical distortions in the estimated

watermark $\hat{w}$ to obtain a rectified estimated watermark $\hat{w}^{rec}$ for detiling (16) and further processing.

13. The method according to one of the claims 9-12, comprising the steps of
    a) generating (17) a reference watermark w_{ref} using a secret reference watermark key and cross-correlating (18) the reference watermark w_{ref} with the block B for identifying and compensating in the block B translations and/or cropping undergone by the stego image y' and/or
    b) averaging identical neighbouring pixels in case of a previously up-sampled block B.

14. The method according to one of the claims 9-13, comprising the steps of
    a) using a secret block allocation key for extracting a codeword c from the block B and/or
    b) using a secret encryption key for decrypting (20) a codeword c of the digital message b and/or
    c) in case of a digital message b having been encoded with a turbo code, using a BJCR, a log-MAP or a max-log-MAP decoder for soft decoding (21) the digital message b.

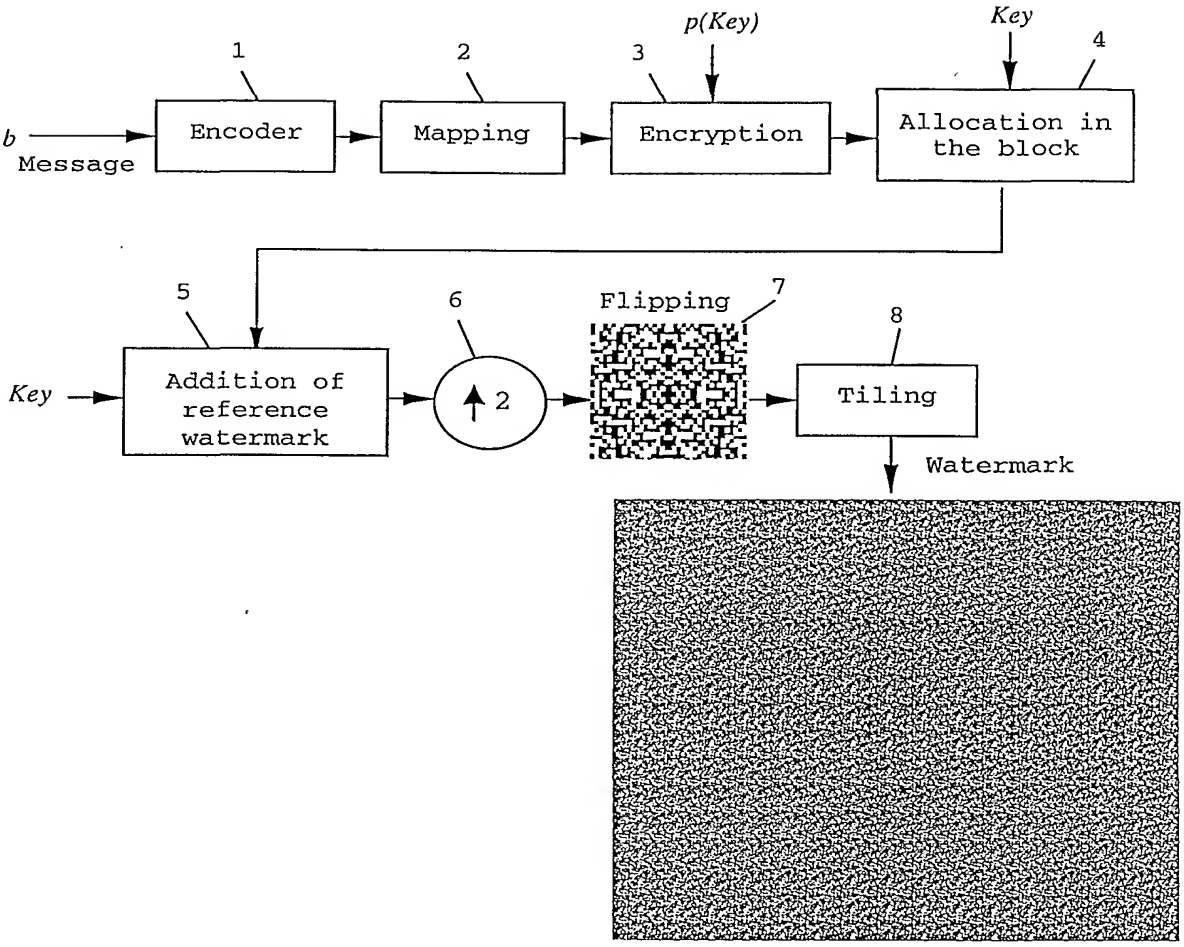

b
Message
1
Encoder
2
Mapping
3
p(Key)
Encryption
Key
4
Allocation in the block
5
Key
Addition of reference watermark
6
2
Flipping
7
8
Tiling
Watermark

Fig.1

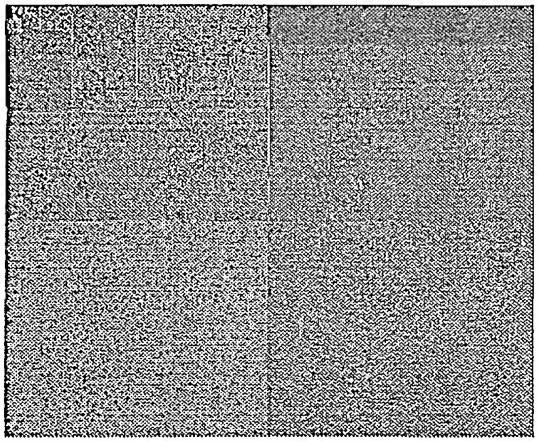


Fig. 2a

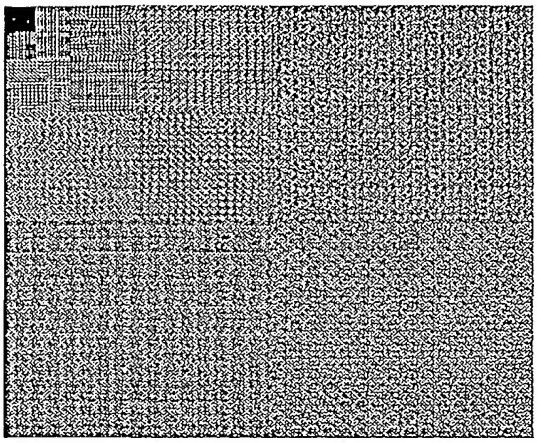


Fig. 2b

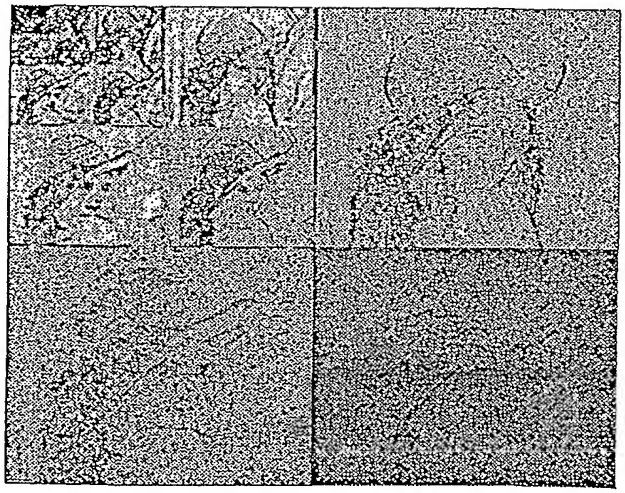


Fig. 2c

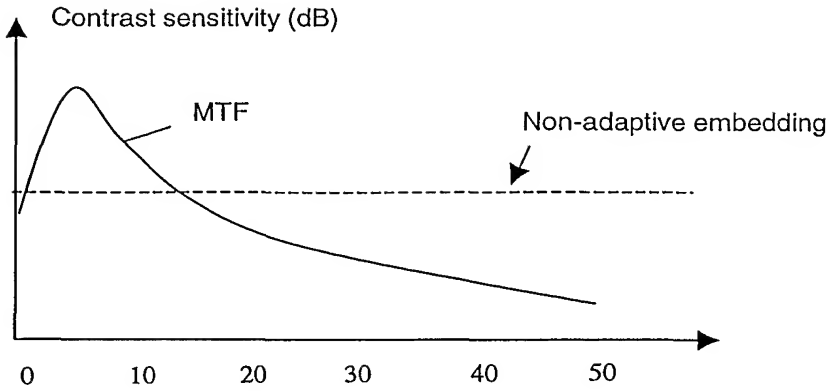


Fig. 3a

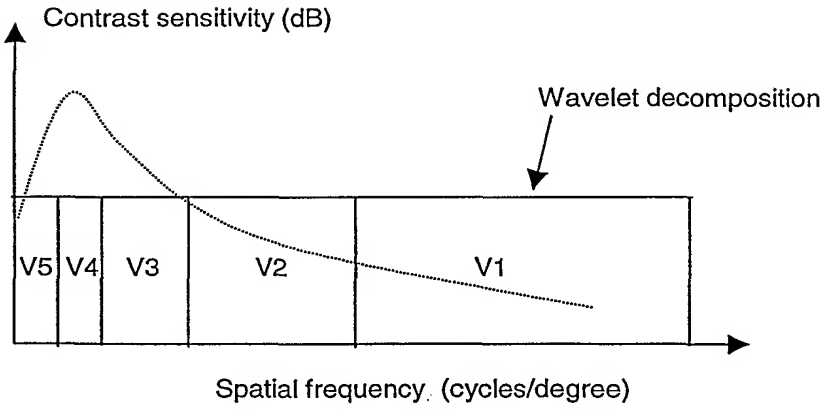


Fig. 3b

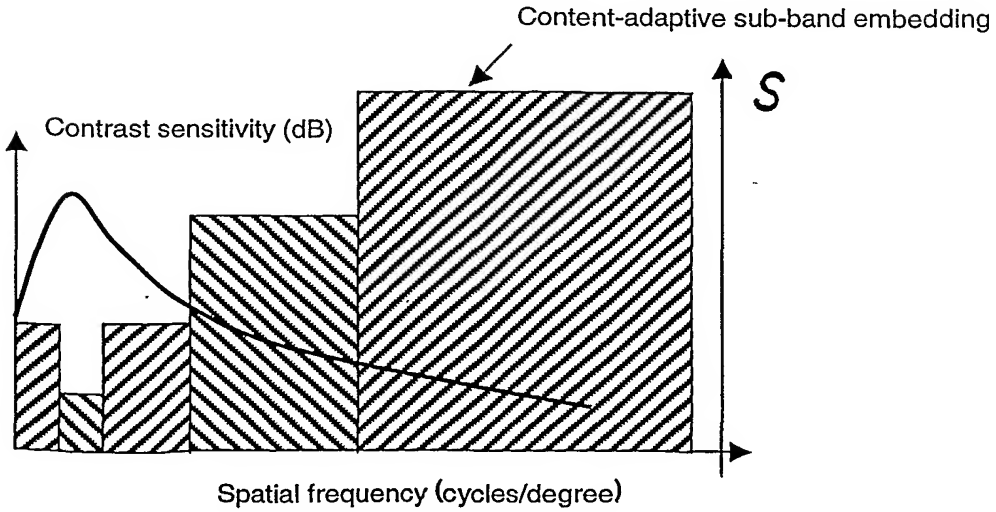


Fig. 3c

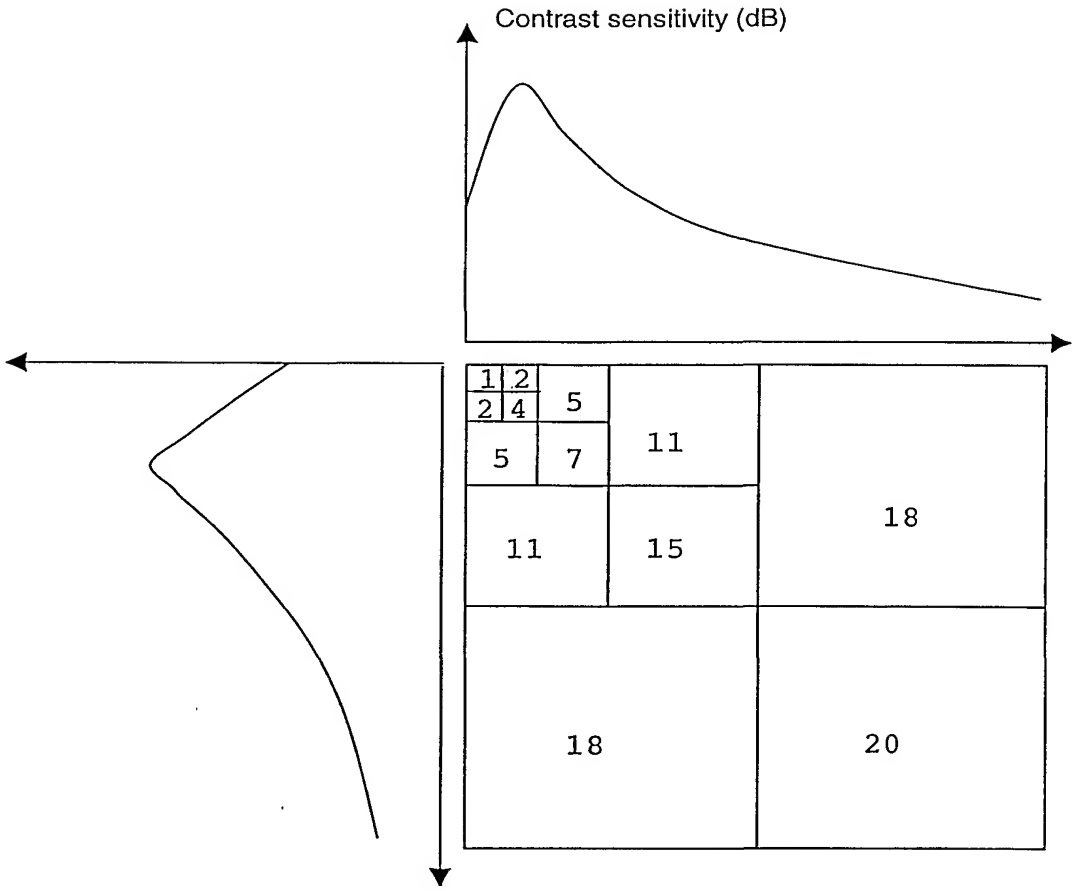

Contrast sensitivity (dB)
1
2
2
4
5
5
7
11
11
15
18
18
20

Fig. 4

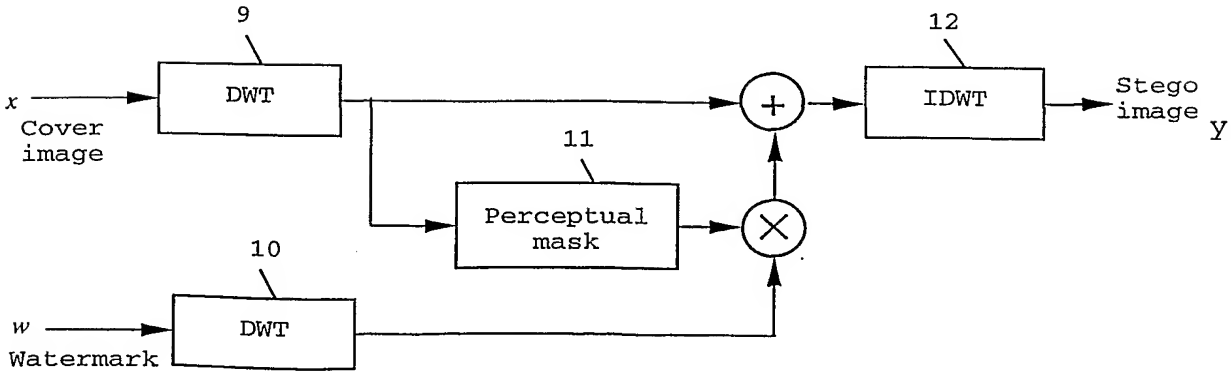

9
DWT
x
Cover image
12
IDWT
Stego image
y
11
Perceptual mask
10
DWT
w
Watermark

Fig. 5

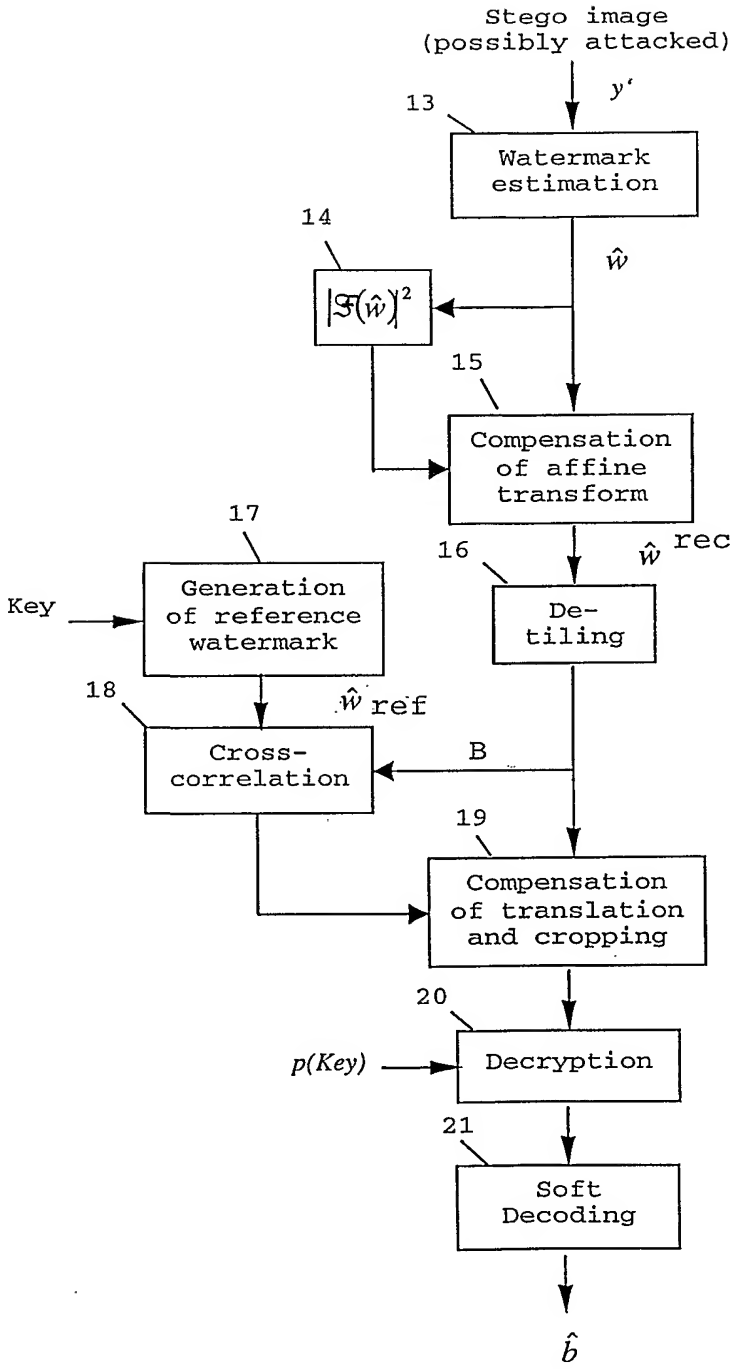

Stego image
(possibly attacked)
y'
13
Watermark estimation
14
$|\mathcal{F}(\hat{w})|^2$
$\hat{w}$
15
Compensation of affine transform
$\hat{w}$ rec
16
De-tiling
17
Key
Generation of reference watermark
18
$\hat{w}$ ref
Cross-correlation
B
19
Compensation of translation and cropping
20
p(Key)
Decryption
21
Soft Decoding
$\hat{b}$

Fig. 6

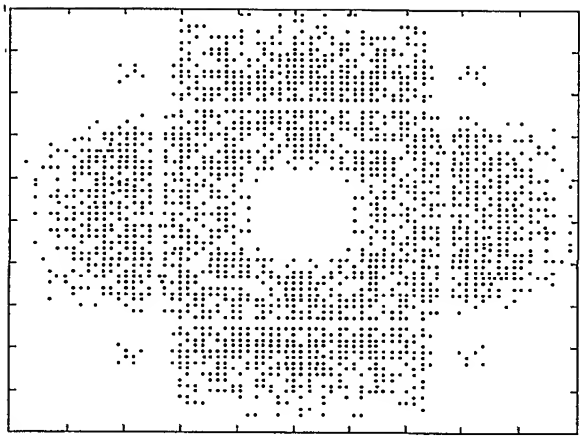


Fig. 7a

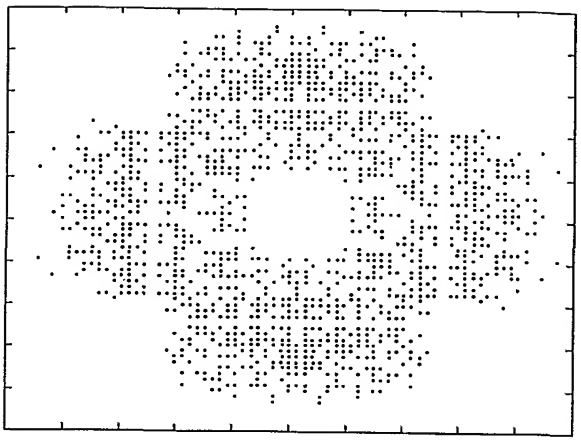


Fig. 7b

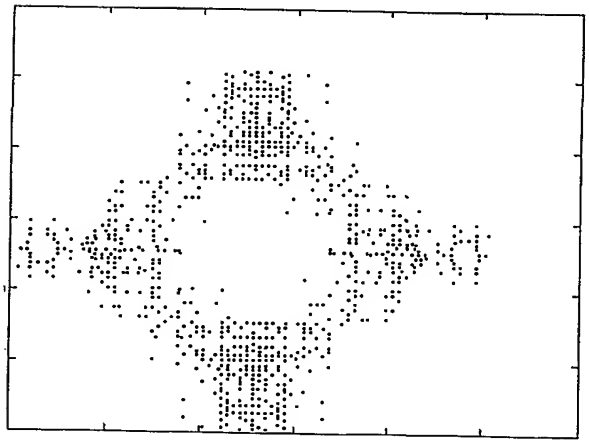


Fig. 7c

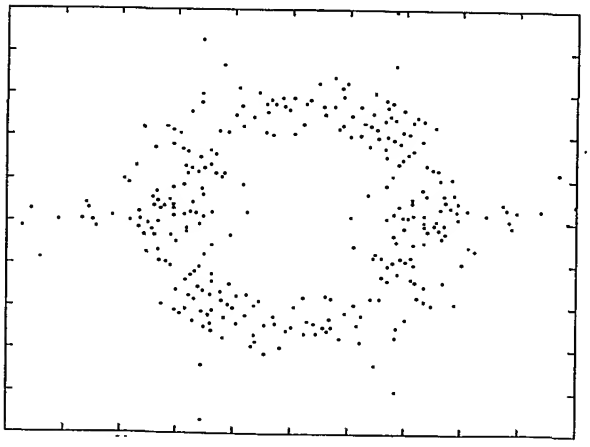


Fig. 7d

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IB 00/01089

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06T1/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, INSPEC, WPI Data, PAJ, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 43736 A (RHOADS GEOFFREY B ;DIGIMARC CORP (US)) 20 November 1997 (1997-11-20) page 55, line 21 - line 29; figure 18	1

[ ] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 April 2001	12/04/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Burgaud, C

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern[redacted]al Application No
PCT/IB 00/01089

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9743736 A	20-11-1997	US 5862260 A	19-01-1999
		US 6122403 A	19-09-2000
		AU 3008697 A	05-12-1997
		EP 1019868 A	19-07-2000